日刊工業新聞
21 ☆ 4月9日・火曜日 2019年（平成31年）

TYPE OF INDUSTRY

科学技術・大学

情報通信研究機構

NICT 先端研究 ⑧④

翻訳バンク 使うほど高精度化

NICTフェロー 隅田 英一郎

日本アイ・ビー・エム、国際電気通信基礎技術研究所を経て情報通信研究機構。アジア太平洋機械翻訳協会会長。日本翻訳連盟理事。内閣府産学官連携功労者表彰総務大臣賞受賞。博士（工学）。

人間でも難しい日本語と英語の間の翻訳には自動化は有用でないと認識されていたが、自動翻訳に人工知能（AI）躍進の要諦である深層学習を導入したところ、ここ数年の間に劇的にその翻訳精度が上がった。英語能力検定のTOEICスコアで900点を越えたともいわれ、既に機械は大半の日本人より高い英語能力を有している。

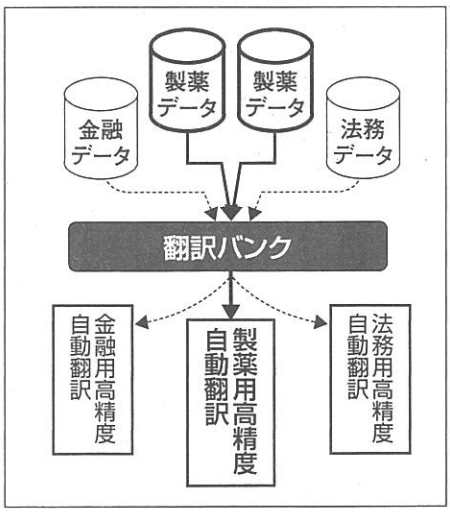
何故この革命的な進化が起こったのだろうか。

AIによる自動翻訳はアルゴリズム（深層学習）と翻訳データ（過去の原文と訳文のこと）で実装されていることから説明できる。

・アルゴリズムについては、研究者が競って新たなアイデアを論文で発表し同時にプログラムを公開するので、世界中にパッと広がり、進歩の限界点は見えていない。

・翻訳データの収集については「翻訳データは量が多ければ多いほど翻訳精度が上がる」ことが多数の実験によって確かめられていて、自動翻訳システムの開発組織間の熾烈な競争下にあり、翻訳データの蓄積はどんどん進んでいる。

この状況を踏まえてNICTは総務省と連携して、多数の組織を糾合し翻訳データを集約するプロジェクト「翻訳バンク」を世界で初めて2017年9月に実稼働させた。①企業・自治体・中央官庁から翻訳データを「翻訳バンク」にご提供いただいて、高精度自動翻訳システムを構築する②そして、この翻訳システムを利用する組織では効率的に新たな翻訳データが生成される③再度、この新データを「翻訳バンク」に提供いただく。このように「翻訳バンク」は自動翻訳をめぐる好循環を実現する。



『翻訳バンク』のコンセプト

日本語にかかわる翻訳データはどれくらいあるのだろうか。翻訳市場の統計から毎年5億文の翻訳データが生産されていると推定される。これらのデータを取り込めれば非常に高精度な自動翻訳システムの誕生は必定である。

既に、翻訳バンクでは、製薬会社から翻訳データの提供を受けて、製薬用の高精度自動翻訳を実現した（図の太い実線の部分に相当）。金融や法務など他の分野のデータも集まり始めており、各分野向けの高精度自動翻訳をどんどん供給して行く。

公的機関に集約するという着眼点が圧倒的なパワーを生み出すことが証明されたわけだ。

今後さらに翻訳データを集積し、翻訳技術の多分野化・高精度化に取り組んでいく。

（火曜日に掲載）